Using zinc finger nuclease technology to generate CRX-reporter human embryonic stem cells as a tool to identify and study the emergence of photoreceptors precursors during pluripotent stem cell differentiation

Journal:	*Stem Cells*
Manuscript ID	SC-15-0423.R1
Wiley - Manuscript Type:	Original Research
Date Submitted by the Author:	n/a
Complete List of Authors:	collin, joseph; Newcastle University, Institute of Genetic medicine Mellough, Carla; University of Newcastle upon Tyne, International Centre for Life , Institute of Human Genetics Dorgau, Birthe; Newcastle University, Institute of Genetic medicine Przyborski, Stefan; University of Durham, UK, SCHOOL OF BIOLOGICAL AND BIOMEDICAL SCIENCE Moreno, Inmaculada; Prince Felipe Research Centre, Lako, Majlinda; University of Newcastle, Institute of Human Genetics;
Keywords:	Differentiation, Embryonic stem cells, Fluorescent protein reporter genes, Gene targeting, Green fluorescent protein, Retina, retinal photoreceptors
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
Collin et al _Suppl. Table 3.doc	

Using zinc finger nuclease technology to generate CRX-reporter human embryonic stem cells as a tool to identify and study the emergence of photoreceptors precursors during pluripotent stem cell differentiation

Joseph Collin [1], Carla Mellough [1], Birthe Dorgau [1], Stefan Przyborski [2], Inmaculada Moreno-Gimeno [3] and Majlinda Lako [1#]

1. Institute of Genetic Medicine, Newcastle University, Newcastle, United Kingdom
2. School of Biological Sciences, Durham University, Durham, United Kingdom
3. Centro de Investigacion Principe Felipe, Valencia, Spain

[#] *to whom correspondence should be addressed:*

Prof. Majlinda Lako

Institute of Genetic Medicine

Newcastle University

Newcastle NE1 B3Z

United Kingdom

Email: majlinda.lako@ncl.ac.uk

Phone: 00 44 191 241 8688

Author contributions: JC designed and performed research, data analysis, figure preparation, manuscript writing and final approval of manuscript; CBM performed research, figure preparation, data analysis and final approval of manuscript; BD performed data analysis, figure preparation and final approval of manuscript; SP performed data analysis and final approval of manuscript; IMG performed data analysis and final approval of manuscript; ML designed and performed research, data analysis, figure preparation, manuscript writing, fund raising and final approval of manuscript.

Keywords: Zinc finger nucleases, CRX, human embryonic stem cells, photoreceptor precursors

Abstract

The purpose of this study was to generate human embryonic stem cell (hESC) lines harbouring the green fluorescent protein (GFP) reporter at the endogenous loci of the Cone-Rod Homeobox (*CRX*) gene, a key transcription factor in retinal development. ZFNs designed to cleave in the 3' UTR of *CRX* were transfected into hESCs along with a donor construct containing homology to the target region, eGFP reporter and a puromycin selection cassette. Following selection, PCR and sequencing analysis of antibiotic resistant clones indicated targeted integration of the reporter cassette at the 3' of the *CRX* gene, generating a CRX-GFP fusion. Further analysis of a clone exhibiting homozygote integration of the GFP reporter was conducted suggesting genomic stability was preserved and no other copies of the targeting cassette were inserted elsewhere within the genome. This clone was selected for differentiation towards the retinal lineage. Immunocytochemistry of sections obtained from embryoid bodies and quantitative RT-PCR of GFP positive and negative subpopulations purified by fluorescence activated cell sorting during the differentiation indicated a significant correlation between GFP and endogenous CRX expression. Furthermore, GFP expression was found in photoreceptor precursors emerging during hESC differentiation, but not in the retinal pigmented epithelium (RPE), retinal ganglion cells or neurons of the developing inner nuclear layer. Together our data demonstrate the successful application of ZFN technology to generate CRX-GFP labelled hESC lines, which can be used to study and isolate photoreceptor precursors during hESC differentiation.

Introduction

Zinc finger nucleases (ZFNs) are designer nucleases which can be engineered to target a specific DNA sequence, offering huge potential for genetically modifying cells with complex genomes, such as mammalian cells [1]. ZFNs are comprised of a DNA binding domain of zinc finger protein motifs (N terminal) fused to the *FokI* endonuclease domain (C terminal), a DNA cleaving domain which operates upon dimerisation [1]. ZFNs are thus designed to work as a pair; upon binding to target sites on opposing strands, they act as a heterodimer and cleave both strands of the DNA. Once the ZFN pair creates a DNA double strand break (DSB) the cell´s inherent DNA repair process is stimulated. In the absence of a repair template, up to 20% of cells can be inaccurately repaired via non homologous end joining (NHEJ), resulting in the imprecise deletion or insertion of bases. In the presence of a donor template containing regions of homology to the target region, homologous recombination can occur resulting in the faithful copying of a template into the endogenous loci, enabling the incorporation of exogenous sequences inserted between two “arms” or regions of homology. Whilst only transient expression of ZFNs is required over a brief period of *in vitro* culture, the resulting genetic manipulation is present for the life of the cell, avoiding the need for continued expression of a foreign transgene. The potential for gene targeting and editing in various complex genomes is therefore considerable and has been realised in multiple organisms [1].

One important application is genetic modification of human pluripotent stem cells. Gene targeting of mouse embryonic stem cells using classical homologous recombination-based methods has been widely used, however such conventional gene targeting approaches are not easily transferred to human embryonic stem cells (hESCs), mainly due to a substantial reduction in efficiency (~10^{-6}, and [2]). However, seminal reports have shown that ZFN technology can be used to carry out precise genome modification of hESCs and human induced pluripotent stem (hiPSC) cells with greater efficiency [3,4]. One application of ZFN technology is to aid the insertion of reporter genes into a specific loci of a genome, rather than relying on the transfection of promoter-driven reporter genes, which often integrate at undetermined sites in the genome that are subject to silencing [5]. By integrating into the endogenous loci the reporter gene is under the same transcriptional control as the gene itself and thus can give a more accurate representation of a genes

activity, as well as protein localisation in the case of reporter fusions. In hESCs, reporters have been used to assess the activity of pluripotency or differentiation markers, as the modification created in the hESCs will be present in the differentiating cells arising from these stem cells. Because the reporter gene should be expressed at a distinct differentiation stage, it should then be possible to use FACS to capture and purify stage-specific cells during tissue morphogenesis. The captured cells can then be characterized in order to define potential stage-specific markers, thus enabling basic biological studies of differentiation and the isolation of progenitor cells derived from human pluripotent stem cells that has not previously been possible. This ability is of particular importance for the isolation of hESC- and hiPSC-derived retinal progenitor cells for which unique cell surface markers are lacking, and for which there is a pressing demand for basic biological studies and cell replacement applications [6].

To date, reporter-based studies for the selection of retinal progenitor cells have been very successful in the mouse [7,8] and this work has shown that the ontogenetic stage of the donor cell is a critical factor for its integration into an adult retina. These studies have highlighted that cells which express *Nrl* (neural retina leucine zipper gene) during early postnatal development display a rod phenotype upon transplantation, while those isolated during the embryonic peak of cone genesis and which express *Crx* (cone-rod homeobox gene), display a cone phenotype following transplantation with a shift towards the rod phenotype if isolated from postnatal retina [7,8]. hESC and hiPSC differentiation towards retinal lineages has undergone substantial advances with the advent of 3D strategies [9-12]; however the same momentum has not been reached with regard to the successful integration of hESC- and hiPSC-derived retinal progenitors transplanted into the adult retina. This could in part be due to the lack of available cell surface markers that can be used to purify desired cell types prior to transplantation, an incomplete understanding of the optimal ontogenetic stage at which to capture differentiating hESCs and hiPSCs to facilitate their functional integration following transplantation, or the inability of current *in vitro* culture conditions to generate progenitor cells that are developmentally equivalent to the *Nrl*- and *Crx*-expressing cells isolated from postnatal murine retina. To begin to answer some of these questions, we set out to generate hESC lines that harbour reporter genes under the control of key transcription factors known to play an important role in retinal

photoreceptor cell commitment. We chose *CRX* as a candidate for the following reasons: (***i***) *CRX* expression is localized to post-mitotic photoreceptor precursors prior to the development of outer segments [13]; (***ii***) clear evidence from animal studies that *CRX* plays an essential role in cone and rod genesis [14]; and (***iii***) existing evidence that the gene's cis regulatory regions can direct reporter gene expression in photoreceptor cells in transgenic studies [15]. Our studies reported herein show that ZFNs can accurately target the *CRX* gene and enable the introduction of a GFP reporter at the 3' terminus of this gene. Furthermore, expression of the GFP reporter mimics the expression of endogenous CRX during the differentiation of hESCs towards a retinal lineage, enabling the fluorescent labelling of CRX-expressing cells with GFP during the differentiation process. Immunocytochemistry with various retinal markers up to day 90 of hESC differentiation indicates that *CRX* expression is observed in hESC-derived photoreceptor precursors which have exited the cell cycle and lack the expression of well-established markers characterising the early eye field, retinal ganglion cells and the inner nuclear layer. Together, these data suggest that CRX-GFP labelled hESC lines created using ZFNs provide an excellent tool with which to study retinal development *in vitro* and photoreceptor precursors at various stages of differentiation, which can be utilised to help define the integration potential of hESC-derived photoreceptor precursors in the intact and diseased retina.

Materials and Methods

hESC culture and differentiation

The H9 hESC line from WiCell Inc. was used in this study. Expansion of hESCs was performed on feeder cells as previously described [16]. The differentiation of H9 CRX-GFP reporter lines towards retinal lineages was performed at least three times under three dimensional (3D) culture conditions, using bacteriological Petri dishes (BD Biosciences) and IGF-1 supplemented media as described in our recent publication [12].

Preparation of ZFN pair, donor construct and nucleofection

The ZFN pair used was designed, produced and activity tested in K562 cells by Sigma-Aldrich (**Suppl. Table 1 and Suppl. Figure 1A**). The activity in hESCs was tested following nucleofection of the ZFN pair mRNA (**Suppl. Figure 1B**). H9 cells were dissociated with Accutase (Life Technologies); $8x10^5$ cells then underwent nucleofection with 2µg of each ZFN mRNA using the Human Stem Cell Nucleofector Kit 1 (Lonza) with a Nuclefector 2b device (Lonza). Following nucleofection cells were cultured as before, with the addition of Y-27632 for 24 hours, and genomic DNA extracted (Quick-gDNA MiniPrep, Zymo Research) 48 hours later. The frequency of cleavage products was then assessed using the SURVEYOR (CEL-I) Mutation Detection Kit (Transgenomic), following the manufacturer's instructions) **Suppl. Figure 1B)**. Briefly, PCR of the region harbouring the target site of the ZFN pair was performed from the genomic DNA (for primers see **Suppl. Table 1**), the PCR products were denatured and then cooled to anneal oligonucleotides, followed by incubation with the CEL-I enzyme at 42°C for 40 minutes. The digests were then run on a 10% PAGE-TBE gel and the fraction of cleaved products assessed.

The design of the donor construct was based on a published sequence by Hockemeyer *et al*. [3], previously used for insertion of a GFP reporter, and included 5' and 3' *CRX* homology arms, eGFP and a PGK-Puro-pA selection cassette (**Figure 1A,** and full sequence in **Suppl. Figure 2**). The construct was synthesised [and cloned into pBS II SK(+)] by Eurofins Genomics. The donor DNA was prepared for transfection by using the QIAGEN Plasmid Plus Maxi kit, followed by linearisation with Psc*I* and gel purification (QIAquick Gel

Extraction Kit, QIAGEN). Nucleofections were performed as before with 2.5µg donor DNA and 2µg of each ZFN mRNA for $8x10^5$ H9 cells. 48 hours post-nucleofection the media was supplemented with 0.5ug/ml Puromycin for 48 hours. Emergent colonies were isolated and expanded.

Pluripotency assays

The expression of pluripotency markers in H9 hESCs harbouring CRX-GFP was determined using the PSC Immunocytochemistry Kit (Life Technologies), following manufacturer's guidelines. Briefly, colonies were fixed in 4% paraformaldehyde, permeabilised with 1% Saponin, blocked with 3% bovine serum albumin (BSA), incubated with OCT4 and SSEA4 antibodies (see **Suppl. Table 3** for details) for 3 hours, washed in PBS, incubated with appropriate secondary antibodies for 1 hour, washed, incubated with DAPI and imaged on a Nikon A1R confocal microscope. At least 10 colonies from each clone were assessed for pluripotency.

Pluripotency of the H9 CRX-GFP clones was then tested by spontaneous differentiation and assessment of the expression of germ layer markers identifying ectoderm (TUJ1), endoderm (AFP) and mesoderm (SMA). Colonies were dissociated by collagenase IV (Life Technologies) and allowed to form embryoid bodies (EBs) in suspension in low-attachment plates containing a general differentiation medium (KO-DMEM, 20% foetal bovine serum, GlutaMAX, 1% NEAA & 1% Pen-Strep). Medium was changed daily for 7 days, EBs were then transferred onto gelatin-coated chamber slides and cultured for a further 7 days. Presence for markers of the germ layers were then assessed using the 3-Germ Layer Immunocytochemistry Kit (Life Technologies) following manufacturers guidelines. Briefly, colonies were fixed in 4% paraformaldehyde, permeabilised with 1% Saponin, blocked with 3% BSA, incubated with TUJ1, AFP and SMA antibodies (see **Suppl. Table 3** for details) for 3 hours, washed in PBS, incubated with appropriate secondary antibodies for 1 hour, washed, incubated with DAPI and imaged on a Nikon A1R confocal microscope. At least five EBs from each clone were assessed for markers of the three germ layers.

PCR, sequencing and copy number analysis

To detect the presence or absence of correctly targeted insertion of the GFP reporter cassette, PCR was performed using GoTaq Long PCR Master Mix (Promega), primers flanking the 5' and 3' *CRX* homology arm sequences (see **Suppl. Table 1**) and genomic DNA extracted (Quick-gDNA MiniPrep, Zymo Research) from puromycin-resistant colonies. PCR products were resolved on a 1% agarose-TAE gel.Sequencing was performed by Eurofins Genomics on genomic DNA extracted from puromycin-resistant colonies and using primers flanking and residing across the donor construct (for primer details see **Suppl. Table 1**).

Copy number analysis of GFP puromycin-resistant clone 1 genomic DNA was carried out using a qPCR approach with TaqMan Copy Number Assays and Genotyping Master Mix (Life Technologies) following manufacturers guidelines. Briefly, a duplex qPCR reaction was performed with an RNase P reference and GFP assay (see **Suppl. Table 1** for details) to quantify the copy number of GFP using clone 1 genomic DNA alongside untreated wild-type H9 hESC genomic DNA. 4 technical replicates were performed for each sample. Reactions were run on a 7500 Fast Real-Time PCR System (Life Technologies) and data analysed in Copy Caller software (Life Technologies).

Investigation of off-site ZFN cleavage

For the prediction of potential off-target binding sites for the 3' CRX ZFN pair the PROGNOS online tool [17] was used. PCR using the primers designed with this tool was then performed (for primer details see **Suppl. Table 2**). PCR products were sequenced by Eurofins Genomics (for sequences see **Suppl. Table 2**). Sequences were subject to BLAST analysis to confirm the presence of insertions or deletions around the predicted off-target site.

Immunocytochemistry

EBs resulting from the directed differentiation of hESCs towards a retinal phenotype were collected at days 30, 60 and 90 of differentiation and immunocytochemistry analysis performed on cryostat sections as previously described [16]. A panel of retinal antibodies listed in **Suppl. Table 3** were used for this analysis. At least five EBs from each differentiation were sampled for immunofluorescent histochemistry. Images were obtained using a Zeiss Axio Imager.Z1 microscope with ApoTome.2 accessory equipment and AxioVision software.

Flow activated cell sorting and quantitative RT-PCR

EBs were disassociated into single cells using 0.05% Trypsin for 30 minutes. Trypsin was inactivated with serum-containing media and the cells washed with PBS. Cells were sorted on a FACS Aria (BD) into GFP positive and negative fractions. RNA was extracted from each fraction (Quick-RNA MicroPrep, Zymo Research), cDNA synthesised (High Capacity cDNA Reverse Transcription Kit, Life Technologies), preamplified (TaqMan PreAmp Master Mix, Life Technologies) and qPCR performed (TaqMan Gene Expression Master Mix, Life Technologies) using a QuantStudio7 system and software (Life Technologies). For TaqMan assays used see **Suppl. Table 1**. Statistical analysis was performed using GraphPad Prism software (Version 6.05) with statistical significance tested using Student's *t*-test.

Karyotype analysis

Karyotypes were determined by standard G-banding procedure as described previously [18]. At least 20 metaphases were analysed for each sample.

Teratoma formation

5-10 x 10^5 hESC were injected subcutaneously into the right flank of in adult SCID mice and maintained for 8-12 weeks [18]. All cells were co-transplanted with 50µl Matrigel (BD Biosciences) to enhance teratoma formation. Two to three animals were injected for each clone. After 8-12 weeks mice were sacrificed, tissues were dissected, fixed in Bouins

overnight, processed and sectioned according to standard procedures and counterstained with either haematoxylin and eosin or Massons Trichrome stain. Sections (5-8μm) were examined using bright field light microscopy and photographed as appropriate.

Results

1. Generation and characterisation of *CRX-GFP* targeted hESC clones

We chose to introduce the GFP reporter at the 3' end of the *CRX* gene in order to avoid disruption of the coding sequence, preventing the loss of *CRX* function and maximising the maintenance of protein localisation and function. For this reason we mutated the stop codon of *CRX* introduced in the targeting cassette, which was followed by the eGFP sequence (**Figure 1A** and **Suppl. Figure 2**), thus resulting in a CRX-GFP fusion upon ZFN pair cutting in the 3'UTR of *CRX* and insertion of the targeting cassette via homologous recombination.

Eighty two colonies arose from the initial puromycin-treated culture following nucleofection. Of these, 24 were picked and expanded. Three of these clones were characterised for the correct integration of the targeting cassette. PCR across the integration site using primers positioned in the genomic DNA and flanking the 5' and 3' *CRX* homologous arms of the donor cassette was performed. This analysis indicated that one of the three selected clones, 'clone 1', harboured a homozygous integration and the other two harboured heterozygous integrations (**Figure 1B**). Sequencing across the integration site was also performed, including the CRX-eGFP fusion border and the homology arm-endogenous genome border confirming correct targeted integration at the 3' end of the *CRX* gene in each clone (**Figure 1A**). All three clones were characterised for karyotypic stability (**Suppl. Figure 3A**) and maintenance of a pluripotent phenotype using *in vitro* assays both at the proliferative stage (**Suppl. Figure 3B**) and following differentiation towards all three germ layers (**Suppl. Figure 3C**), in addition to *in vivo* teratoma formation assays (**Suppl. Figure 3D**). As we set out to assess whether ZFN technology could be used to precisely tag a gene of interest with minimal disruption of the gene environment in order to produce an accurate reporter of gene expression then the most stringent test for this was to use a clone where

homozygous tagging of the gene had occurred and for this reason we continued our analyses with clone 1. Further analysis to assess the genomic fidelity of clone 1 following ZFN-mediated tagging was then conducted. Quantitative PCR based copy number analysis of GFP revealed the presence of two copies of GFP (**Figure 1C**), corroborating the PCR and sequencing data and suggesting that no additional copies of GFP were present in the genome due to non-targeted insertion of the donor cassette. In addition, the 15 most likely genomic sites for targeting by the ZFN pair were predicted (**Suppl. Table 2**) using the PROGNOS online tool [17]. Using the primers generated from this analysis genomic DNA from clone 1 was amplified and sequenced. This analysis indicated no insertions, deletions or integration at predicted off-target sites. Together our data suggest that ZFN targeting at the 3' end of *CRX* in hESCs does not affect the maintenance of genomic stability or pluripotency.

2. GFP reporter mimics the expression of endogenous CRX during retinal differentiation of hESC targeted clones

The CRX-GFP targeted hESC clones were subjected to our 3D differentiation protocol described in Ref. [12]. This protocol utilises a single growth factor (IGF-1) and results in formation of laminated neural retina comprising a range of well differentiated retinal phenotypes [12]. This differentiation method relies on the mechanical transfer of hESC colonies to suspension conditions, forming 3D EBs which give rise to laminated neural retinal tissue displaying apically positioned photoreceptors; however a minority of EBs show internal cellular rosettes exhibiting a central hollow lumen with CRX positive cells found towards the internal surface. In accordance with this, we were able to detect GFP positive cells residing in the apical aspect of developing retinal tissue (**Figure 2A-D**) as well as the internal aspect of retinal rosettes (**Figure 2E-L**) during the 90 day differentiation protocol.

On days 30, 60 and 90 of differentiation, floating EBs were collected and subjected to immunocytochemical analysis using antibodies directed against GFP and CRX (**Figure 2A-L**). This analysis showed an identical correlation between GFP and endogenous *CRX* expression throughout the time-course of differentiation. Furthermore, the detection of GFP expression could be carried out without the need for an anti-GFP antibody (**Figure 2M**). Together our data suggest that a GFP reporter introduced at the 3' terminal of *CRX* does not

adversely affect the expression of *CRX* and can accurately mimic the expression of endogenous *CRX*.

3. CRX/GFP expression is confined to photoreceptor precursors during hESC differentiation

Throughout the differentiation period immunocytochemical analysis was performed using markers characterising the early stages of eye field formation, emergence of photoreceptor precursors, mature photoreceptors and other retinal phenotypes. Co-immunostaining with an anti-GFP antibody at day 60 of differentiation revealed no co-expression with early eye field markers (**Suppl. Figure 4**), indicating that CRX is not expressed with RAX6 or PAX6 during the early stages of optic cup formation. As differentiation proceeded, abundant RECOVERIN expression was observed at days 60 and 90 (**Figure 3A, 3B** and **4A**) in the outer nuclear layer of the developing optic cup as reported in our recent publication [12]. Interestingly, all RECOVERIN-positive cells co-expressed CRX, while only 70% of CRX-expressing cells were double-labelled with RECOVERIN (data not shown). A similar expression pattern was observed in the developing human retina at 16 and 18 weeks of gestation (**Suppl. Figure 5A-D**). Furthermore, there was no overlap in the expression of GFP with Ki67, suggesting that the CRX-expressing cells were non-proliferative (**Figure 3G** and **4F**). In addition, no CRX expression was observed in the emerging RPE layer (data not shown) over the differentiation period.

Expression of cone photoreceptor markers (OPN1SW and OPN1MW+OPN1LW) was tested at days 60 and 90 of differentiation and a punctate cytoplasmic expression pattern was observed (**Figures 3C-D** and **Figures 4B-C**), perhaps indicative of very early stages of cone photoreceptor genesis which lacks the typical plasma membrane localisation of Opsins reported at later stages of differentiation [11]. Similar findings were reported by Kaewkhaw et al. [19] who noted minimal short wave opsin and no rhodopsin or medium and long wave opsin expression at day 90 of hESC differentiation. Double-labelling of tissue with GFP and cone photoreceptor markers (Opsin Blue/OPN1SW and Opsin Red+Green/OPN1MW+OPN1LW, **Figures 3C-D** and **Figures 4B, C**) or Calbindin (expressed in several cell types including cone photoreceptors, **Figure 3E**) indicated no obvious co-expression, suggesting a lack of CRX expression in developing cone photoreceptors emerging within hESC derived optic cups on days 60 and 90 of hESC differentiation.

No co-expression was observed for CRX with PKCα (expressed in Rod ON Bipolar Cells, **Figure 4D**) or HuC/D (expressed in amacrine and ganglion cells, **Figure 3F** and **4E**) or, suggesting a lack of CRX expression in retinal ganglion cells and developing inner nuclear layer retinal neurones. These results were further corroborated by fluorescence-activated cell sorting combined with quantitative RT-PCR (**Figures 4G-I**) where significantly higher expression of retinal photoreceptor precursor markers *(CRX*, *RECOVERIN* and *CONE ARRESTIN*) was found in the GFP positive fraction (**Figure 4H**). In contrast, higher expression of markers of post mitotic rod precursors (*NRL*) (**Figure 4H**), mature rods (*RHO*) and cone photoreceptors (*OPN1SW, OPN1MW, OPN1LW*) were found in the GFP negative fraction (**Figure 4I**). To evaluate the developmental stage of CRX-GFP positive cells, we performed qRT-PCR analysis of the same markers in the developing human retina obtained at 18 weeks of gestation. *CRX* was expressed at similar levels, whilst *RECOVERIN* was expressed at a higher level in the developing retina when compared to the CRX-GFP positive population (**Figure 4H**). Conversely, cone arrestin was expressed at higher levels in the CRX-GFP positive population (**Figure 4H**). Together these data suggest that the CRX-GFP positive population shows similarities to the 18 week foetal retina with high expression of photoreceptor markers but due to the greater cone expression may be at a different developmental stage or consist of an enriched precursor population, however to fully ascertain this further investigation with expression studies at other developmental and differentiation time points would be required and are beyond the scope of this paper.

In summary, our data suggest that *CRX* expression is confined to two cell types during the first 90 days of hESC differentiation, namely *RECOVERIN*-expressing photoreceptor precursors situated in the developing outer nuclear layer of the optic cup and in a subpopulation of retinal progenitors that appear to be non-proliferative but do not yet express any of the markers characterising the mature retinal phenotypes that we have tested herein. Detection of *CRX* expression as well as genes activated by *CRX* (for example *CONE ARRESTIN)* indicates that CRX remains functional following the ZFN-mediated tagging with GFP.

Discussion

Reporter lines that facilitate the detection of emerging photoreceptors are a useful tool for monitoring and improving hESC/hiPSC differentiation protocols towards retinal lineages and also enable the selection of specific cell types, analysis of their molecular and cell surface profile and engraftment capacity. Using plasmid and lentiviral based approaches, a few reporter hESC and hiPSC lines have been generated in order to monitor the emergence of retinal ganglion like cells [20] and photoreceptors [21]. Both of these approaches can suffer from random integration into the genome and subsequent silencing of reporter cassettes [5]. Insertion of promoter-reporter constructs into safe loci (for example, AAVS1) has also been suggested as a potential way of circumventing the silencing of the reporter cassettes; however this method still suffers from the limitations imposed by the selected promoter region being inserted into the cassette, which may not fully replicate the regulatory region of the endogenous gene. To bypass these issues, we have applied ZFN technology combined with the inherent repair of DNA DSBs by homologous recombination to generate a hESC line which harbours the GFP reporter as a 3' terminal fusion of the endogenous *CRX* gene, a well described marker of post mitotic photoreceptor precursors. The reason for targeting 3' of *CRX* was based on previous findings in the mouse system which indicated that the 5'UTR (up to 12 Kb) was not sufficient to direct high level reporter gene expression in photoreceptor cells compared to a mixture of 5' and 3' UTR regions in murine transgenic lines [15]. Generating a 3' reporter fusion presents several advantages, such as maintenance of the full-length gene sequence with less likelihood of disruption of the coding sequence whilst maximising the similarity of reporter expression to that of the endogenous gene, however it also presents technical challenges related to the presence of AT-rich sequences commonly found at the 3' UTR [22], which are notoriously difficult to clone (and hence to construct the integrating cassette) and are also less preferable for ZFN targeting when compared to GC-rich regions [23, 24]. This potential for lower integration efficiency did not pose a problem for our study as we were able to isolate 82 clones and further characterise three CRX-GFP targeted hESC clones (one homozygous and two heterozygous), which met the full assessment criteria for pluripotency, indicating that CRX targeting is compatible with the maintenance of pluripotency. Most importantly, no genomic abnormalities were observed after copy number analysis and sequencing of the

top 15 most likely cutting sites, suggesting that the CRX-GFP targeted clones generated in this study provide a *bona fide* tool for hESC retinal differentiation studies.

Since *CRX* is not expressed at the pluripotent stem cell stage we subjected our targeted clones to our recently described 3D retinal differentiation protocol [12] and performed immunocytochemistry with antibodies against CRX and GFP during both early and late stages of differentiation. The anti-GFP antibody was employed for co-immunocytochemistry with antibodies against cellular markers due to the observation of reduced GFP fluorescence following processing for immunocytochemistry and to obtain comparable levels of fluorescence intensity between both antibodies allowing optimal image clarity when performing the co-staining. With this method we were able to show perfect correlation between the expression and localisation of CRX and GFP, suggesting that the GFP reporter accurately mimics the expression of endogenous *CRX* during hESC differentiation. Importantly, we were also able to detect GFP expression without the need for immunocytochemistry indicating that GFP confers sufficient reporter activity to monitor the emergence of *CRX* expressing cells during hESC differentiation. The ability to detect CRX expression also indicated that expression of CRX had not been adversely hampered by ZFN-mediated tagging with GFP. Furthermore, CRX protein has been shown to be required for the transactivation of its own gene expression and its maintenance [25]; henceforth, continued detection of CRX mRNA and protein expression throughout our differentiation time course further reinforces the notion that ZFN-mediated GFP tagging has not interfered with CRX function.

Expression studies in human foetal retina have suggested that *CRX* is expressed in photoreceptors and in the inner nuclear layer [26] and mouse transgenic studies have also shown *CRX* expression in bipolar cells in addition to photoreceptors [15, 25]. To investigate the expression of *CRX* during hESC differentiation we performed immunocytochemistry with retinal precursor and retinal lineage markers. We further validated this analysis by performing quantitative RT-PCR of GFP positive and negative populations isolated by FACS. We found that the GFP positive population displayed increased expression of photoreceptor markers (*CRX, RECOVERIN* and *CONE ARRESTIN*) when compared to the GFP negative population or unsorted cells. In addition to transactivating its own expression, CRX protein has been shown to be important for activating and maintaining the expression of several

key photoreceptor genes [25, 27, 28, 29, 30]. For example, the *CONE ARRESTIN* gene harbours several CRX binding sites in its promoter region which are important for its activation [31, 32, 33], hence expression of *CONE ARRESTIN* in the CRX-GFP positive population indicates that the CRX protein remains functional and able to activate downstream targets after ZFN-mediated GFP tagging.

It was interesting to observe that the expression of *RHODOPSIN*, a mature rod photoreceptor marker, was greater in the GFP negative subpopulation, suggesting confined expression of the GFP reporter to photoreceptor precursors emerging during 90 days of hESC differentiation. While we could detect expression of cone opsins in a punctate pattern that may be typical of an immature phenotype, we were unable to observe mature rod markers by immunocytochemistry, which indicates either a very low frequency of rod differentiation at day 90, or that the ontogenic stage of day 90 hESC-derived retinal tissue precedes the developmental peak of rod genesis, akin to the pattern of retinal cell emergence during human development. These data however should be interpreted with caution, as the expression of opsins in our study was low or of an immature nature. These findings are corroborated by a recent publication [19] and suggest that longer differentiation experiments are required to investigate CRX reporter expression in photoreceptors that display an advanced stage of morphological features and electrophysiological function. When compared to developmental retina at 18 weeks of gestation, higher *CONE ARRESTIN* expression was observed in the CRX-GFP positive population, perhaps indicating a more advanced stage of cone genesis in the CRX positive population isolated from hESC differentiation, though this needs to be further investigated using comparative analysis with retina obtained at additional developmental periods. Intriguingly NRL expression is observed in the foetal retina but is absent from the CRX-GFP positive population. This together with the expression of *CONE ARRESTIN* may mark these cells as cone-specific progenitors; however the heterogeneous nature of the hESC differentiation could equally yield populations of varying maturity with some cells beginning to acquire a cone photoreceptor fate whilst other cells are at an earlier post-mitotic photoreceptor progenitor phase and rod genesis is yet to occur.

As outlined by previous studies in the mouse model [8] we also found that CRX expression was a marker of retinal progenitor cells that had exited the cell cycle, as assessed

by lack of Ki67 expression in the GFP positive subpopulation. We also found a small percentage of CRX positive cells that did not co-express RECOVERIN both in hESC-derived laminated retina and in human foetal retina. It is already known that in the developing retina progenitor cells undergo interkinetic nuclear migration, where the nucleus oscillates in an apical to basal fashion throughout the full thickness of the retinal neuroepithelium in synchrony with the cell cycle. DNA duplication occurs towards the basal surface whilst mitosis occurs at the apical aspect. Some of the CRX-GFP positive cells lacking *RECOVERIN* expression were situated towards the apical surface, so it is possible that they move towards this rim to complete mitosis before acquiring RECOVERIN expression and committing to a photoreceptor fate. If this were the case, these cells should also express Ki67. However our analysis indicated that all CRX positive cells lack Ki67 expression leaving open the question of what these cells may be. Earlier studies in mouse transgenic models have shown that CRX expression can also be detected in bipolar cells [15]; however we did not observe any CRX cells to express PKCα, ruling out the subset of bipolar cells characterised by the expression of this marker. Clearly, further studies are required to investigate this CRX population more completely; however the GFP reporter described in this manuscript would also enable laser dissection capture of this subset of CRX positive cells and permit further analysis at the transcriptional level to help determine their identity.

In conclusion we have successfully generated CRX-GFP reporter hESC lines, which can be used to study the molecular and cell surface profile of human photoreceptor precursors and enable their isolation for transplantation studies. We hope that these hESC lines will provide a universally useful tool with which to monitor and improve differentiation protocols, discover useful cell surface markers and develop clinically applicable strategies for the purification of hESC-derived retinal photoreceptor precursors for transplantation.

ACKNOWLEDGEMENTS

The authors are grateful to BBSRC UK (#BB/I02333X/1), ERC (#614620) and Fight for Sight UK (#1870) for funding this work and to the MRC-Wellcome Trust Human Developmental Biology Resource for provision of human developmental tissues (grant number grant# 099175/Z/12/Z). The authors are also grateful to Jerome Evans for performing the karyotype analysis reported in this study.

Note added in proof: Whilst this paper was under revision, Swaroop and colleagues [19] reported similar findings on generation of a CRX-GFP marked hESC line, which utilised a CRX promoter based plasmid to mark the expression of emerging CRX progenitors during the retinal differentiation process.

Conflict of interest

The authors declare no conflict of interest.

References

1. Collin J and Lako M. Concise review: putting a finger on stem cell biology: zinc finger nuclease-driven targeted genetic editing in human pluripotent stem cells. Stem Cells. 2011 Jul;29(7):1021-33
2. Zwaka TP, Thomson JA. Homologous recombination in human embryonic stem cells Nat Biotechnol. 2003 Mar;21(3):319-21.
3. Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKelver RC,et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol. 2009 Sep;27(9):851-7.
4. Zou J, Maeder ML, Mali P, Pruett-Miller SM, Thibodeau-Beganny S, Chou BK, et al. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell. 2009 Jul 2;5(1):97-110.
5. Stewart R, Yang C, Anyfantis G, Przyborski S, Hole N, Strachan T et al. Silencing of the expression of pluripotent driven-reporter genes stably transfected into human pluripotent cells. Regen Med. 2008 Jul;3(4):505-22.
6. Mellough CB, Collin, J, Sernagor E, Wride NK, Steel DH and Lako M. Lab generated retina: realizing the dream. Vis Neurosci. 2014 Sep;31(4-5):317-32.
7. MacLaren RE, Pearson RA, MacNeil A, Douglas RH, Salt TE, Akimoto M. Retinal repair by transplantation of photoreceptor precursors. Nature. 2006 Nov 9;444(7116):203-7.

8. Lakowski J, Baron M, Bainbridge J, Barber AC, Pearson RA, Ali RR et al. Cone and rod photoreceptor transplantation in models of the childhood retinopathy Leber congenital amaurosis using flow-sorted Crx-positive donor cells. Hum Mol Genet. 2010 Dec 1;19(23):4545-59.
9. Eiraku M, Takata N, Ishibashi H, Kawada M, Sakakura E, Okuda S et al. Self-organizing optic-cup morphogenesis in three-dimensional culture. Nature. 2011 Apr 7;472(7341):51-6.
10. Nakano T, Ando S, Takata N, Kawada M, Muguruma K, Sekiguchi K et al. Self-formation of optic cups and storable stratified neural retina from human ESCs. Cell Stem Cell. 2012 Jun 14;10(6):771-85.
11. Zhong X, Gutierrez C, Xue T, Hampton C, Vergara MN, Cao LH et al. Generation of three-dimensional retinal tissue with functional photoreceptors from human iPSCs. Nat Commun. 2014 Jun 10;5:4047.
12. Mellough CB, Collin J, Khazim M, White K, Sernagor E, Steel D and Lako M. IGF-1 signalling plays an important role in the formation of three dimensional laminated neural retina and other ocular structures from human embryonic stem cells. Stem Cells. 2015 Aug;33(8):2416-30.
13. Furukawa T, Morrow EM, Cepko CL. Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. Cell. 1997 Nov 14;91(4):531-41.
14. Freund CL, Gregory-Evans CY, Furukawa T, Papaioannou M, Looser J, Ploder L et al. Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. Cell. 1997 Nov 14;91(4):543-53.
15. Samson M, Emerson MM, Cepko CL. Robust marking of photoreceptor cells and pinealocytes with several reporters under control of the Crx gene. Dev Dyn. 2009 Dec;238(12):3218-25.
16. Mellough CB, Sernagor E, Moreno-Gimeno I, Steel DH, Lako M. Efficient stage-specific differentiation of human pluripotent stem cells toward retinal photoreceptor cells. Stem Cells 2012; 30(4):673-86.
17. Fine EJ, et al. An online bioinformatics tool predicts zinc finger and TALE nuclease off-target activity. Nucleic Acids Research 2014; 42(6):e42.
18. Stojkovic M, Lako M, Stojkovic P, Stewart R, Przyborski S, Armstrong L, et al. Derivation of human embryonic stem cells from day-8 blastocysts recovered after three-step in vitro culture. Stem Cells 2004; 22(5): 790-7
19. Kaewkhaw R, Kaya KD, Brooks M, Homma K, Zou J, Chaitankar V, Rao M and Swaroop, A. Transcriptome dynamics of developing photoreceptors in three-dimensional retina cultures recapitulates temporal sequence of human cone and rod differentiation revealing cell surface markers and gene networks. *Stem Cells* 2015 doi: 10.1002/stem.2122.

20. Tanaka T, Yokoi T, Tamalu T, Watanabe S-I, Nishina S and Azuma N. Generation of retinal ganglion cells with functional axons from human induced pluripotent stem cells. Sci Rep. 2015 Feb 10;5:8344.
21. Lamba DA, McUsic A, Hirata RK, Wang PR, Russell D, Reh TA. Generation, purification and transplantation of photoreceptors derived from human induced pluripotent stem cells. PLoS One. 2010 Jan 20;5(1):e8763.
22. Barrett LW, Fletcher S, Wilton SD. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell Mol Life Sci. 2012 Nov;69(21):3613-34.
23. Porteous MH, Caroll D. Gene targeting using zinc finger nucleases. Nat Biotechnol. 2005 Aug;23(8):967-73.
24. Zhang C, Xiao B, Jiang Y, Zhao Y, Li Z, Gao H, Ling Y, Wei J, Li S, Lu M, Su Z, Cui H, Yuan J. Efficient Editing of Malaria Parasite Genome Using the CRISPR/Cas9 System. MBio. 2014 Jul 1;5(4): e01414-14. doi:10.1128/mBio.01414-14.
25. Furukawa A, Koike C, Lippincott P, Cepko CL, Furukawa T. The mouse Crx 5'-upstream transgene sequence directs cell-specific and developmentally regulated expression in retinal photoreceptor cells. J Neurosci. 2002 Mar 1;22(5):1640-7.
26. Bibb LC, Holt JK, Tarttelin EE, Hodges MD, Gregory-Evans K, Rutherford A, Temporal and spatial expression patterns of the CRX transcription factor and its downstream targets. Critical differences during human and mouse eye development. Hum Mol Genet. 2001 Jul 15;10(15):1571-9.
27. Chen S, Wang QL, Nie Z, Sun H, Lennon G, Copeland NG, Gilbert DJ, Jenkins NA, Zack DJ. Crx, a Novel Otx-like Paired-Homeodomain Protein, Binds to and Transactivates Photoreceptor Cell-Specific Genes Neuron. 1997 Nov;19(5):1017-30.
28. Sakuma H, Murakami A, Fujimaki T, Inana G. Isolation and characterization of the human X-arrestin gene. Gene 1998 Dec 11;224(1-2):87-95
29. Furukawa T, Morrow EM, Li T, Davis FC, Cepko CL. Retinopathy and attenuated circadian entrainment in *Crx*-deficient mice. Nat Genet. 1999 Dec;23(4):466-70.
30. Kimura A, Singh D, Wawrousek EF, Kikuchi M, Nakamura M, Shinohara T. Both PCE-1/RX and OTX/CRX Interactions Are Necessary for Photoreceptor-specific Gene Expression. J Biol Chem. 2000 Jan 14;275(2):1152-60.
31. Zhu X, Ma B, Babu S, Murage J, Knox BE, Craft CM. Mouse cone arrestin gene characterization: promoter targets expression to cone photoreceptors. FEBS Lett. 2002 Jul 31;524(1-3):116-22.
32. Pickrell SW, Zhu X, Wang X, Craft CM. Deciphering the Contribution of Known cis-Elements in the Mouse Cone Arrestin Gene to its Cone-Specific Expression. Invest Ophthalmol Vis Sci. 2004 Nov;45(11):3877-84.
33. Fujimaki T, Huang ZY, Kitagawa H, Sakuma H, Murakami A, Kanai A, McLaren MJ, Inana G. Truncation and mutagenesis analysis of the human X-arrestin gene promoter 2004

34. Rapaport DH, Stone J. The area centralis of the retina in the cat and other mammals: focal point for function and development of the visual system. Neuroscience. 1984 Feb;11(2):289-301.

Figure legends

Figure 1: ZFN mediated targeting of GFP to the 3' UTR of *CRX*. A) Schematic diagram of endogenous *CRX* gene structure, targeting cassette, targeting strategy and sequencing data from sections of the insertion site, the endogenous CRX-exogenous eGFP border site and 3' homology arm-endogenous *CRX* border, obtained from three analysed hESC clones; **B**) PCR of three hESC clones with primers spanning the integration site to assess whether the cassette had integrated at the correct site. Depending on size of band produced (upper band of 5067bp with integration, lower band of 2171bp without) and whether the clone has a heterozygous integration (by the presence of the upper and lower bands), or a homozygote integration (a single upper band), such as clone 1 in the first lane. Lanes 2 & 3 show two clones with heterozygote integration, lanes 4-6 are genomic DNA from control untreated H9 hESCs and lane 7, a no template negative PCR control. This is a representative example of at least 3 repeats; **C**) qPCR mediated copy number analysis indicating presence of two copies of GFP (and hence targeting cassette) in CRX-GFP hESC clone 1 (c1) and its absence from wild type H9 hESC genomic DNA, data is presented as the mean ± SEM (n=3).

Figure 2. GFP reporter accurately mimics the expression of endogenous CRX during hESC differentiation. (**A-L**) Immunocytochemistry with antibodies raised against CRX and GFP showing correlation at day 30 (A-D), day 60 (E-H) and day 90 (I-L) of hESC differentiation. **M**) Detection of GFP without antibody using confocal microscopy. Scale bars= 20 µm.

Figure 3. CRX/GFP expression is found in photoreceptor precursors at day 60 of hESC differentiation. Immunocytochemistry with antibodies raised against GFP and Recoverin (**A**

and **B**), OPN1SW (**C**), OPN1MW/LW (**D**), Calbindin (**E**), HuC/D (**F**) and Ki-67 (**G**). Scale bars = 20 µm (A, F), 5 µm (B) and 10 µm (C, D and E).

Figure 4. CRX expression in photoreceptor precursors assessed by immunocytochemistry and qRT-PCR at day 90 of differentiation. Immunocytochemistry with antibodies raised against GFP and Recoverin (**A**), OPN1SW (**B**), OPN1MW/LW (**C**), PKCα (**D**), HuC/D (**E**) and Ki-67 (**F**). Scale bars = 10 µm (E) and 20 µm (A, B, C, D and F); (**G**) Flow activated cell sorting of GFP positive (green) and GFP negative (red) populations; (**H**) and (**I**) qRT-PCR analysis of unsorted, flow activated cell sorted GFP positive (Sorted GFP+) and GFP negative (Sorted GFP-) populations and foetal retina at 18 weeks of gestation (18w Foetal Retina), indicating *CRX*, *RECOVERIN* (*RCVRN*) and *CONE ARRESTIN* (*C-ARR*) expression in the GFP positive population. Expression of short, medium and long wave opsins (*OPN1SW*, *OPN1MW* & *OPN1LW*) and rhodopsin (*RHO*) are also shown. Data is presented as the mean ± SEM (n=3). Significant differences found are marked with an asterisk (*).

Suppl. Figure 1: Surveyor mutation detection assay performed in K562 cells and hESCs to indicate ZFN cutting efficiency. As shown, a higher ZFN cutting efficiency is observed in K562 cells than hESCs (which may be related to lower non homologous end joining repair pathway efficiency observed in hESCs). Upon cleavage of the target site by the ZFN pair inaccurate repair by the non-homologous end joining pathway can occur resulting in insertion or deletion of nucleotides. Following PCR of the target region and cooling of the PCR products duplexes form. In the presence of NHEJ-repaired and wild-type heteroduplexes and the SURVEYOR/CEL-I enzyme cleavage products are produced. The 221 and 150 bp bands represent the cleavage of PCR products generated across the *CRX* 3' UTR target cut site in the presence of an insertion or deletion, in the absence of an insertion or deletion a wild-type PCR product of 371 bp is observed. (**A**) K562 cells; lane 1, PCR of wild-type genomic DNA (No), lanes 2 & 3, PCR of genomic DNA following ZFN pair transfection (ZFN); (**B**) hESC H9 cells; lane 1, PCR of wild-type genomic DNA (No), lane 2, PCR of genomic

DNA following a mock nucleofection without ZFNs (Mock) & lane 3, PCR of genomic DNA following ZFN pair nucleofection (ZFN).

Suppl. Figure 2: The sequence of CRX-GFP targeting construct. The sequences of *CRX* 5' and 3' homology arms, eGFP reporter and puromycin resistance cassette are shown as highlighted in the key.

Suppl. Figure 3: CRX-GFP targeted hESC clones maintain a normal karyotype and exhibit pluripotency. (**A**) Normal karyotype (46 XX) observed in CRX-GFP hESC clone 1; (**B**) Pluripotent marker staining (OCT4 and SSEA4) in CRX-GFP hESC clone 1; (**C**) Three germ layer differentiation capacity assessed by the presence of endodermal cells (marked by AFP staining), ectodermal cells (marked by TUJ1 staining) and mesodermal cells (marked by SMA) in embryoid bodies derived from differentiation of CRX-GFP hESC clone 1; (**D**) Histological analysis of xenograft tumours formed from engrafted CRX-GFP hESC clone 3. The teratomae produced contained tissues representative of endoderm (***a***), mesoderm (***b***) and ectoderm (***c***). Example tissues included: (***a***) structure of primitive intestine showing epithelium (ep), villous (vi), submucosa (sb), and smooth muscle (sm); (***b***) cartilaginous masses (ct); (***c***) neuro-epithelium (ne). Counterstain: Haematoxylin and Eosin. Scale bars: (a-c) 150 microns.

Suppl. Figure 4: CRX is not expressed in early stages of eye cup formation. Immunocytochemistry with antibodies raised against GFP, RAX (**A**) and PAX6 (**B**) (both markers of eyefield development) at day 60 of differentiation. Scale bars= 20 µM.

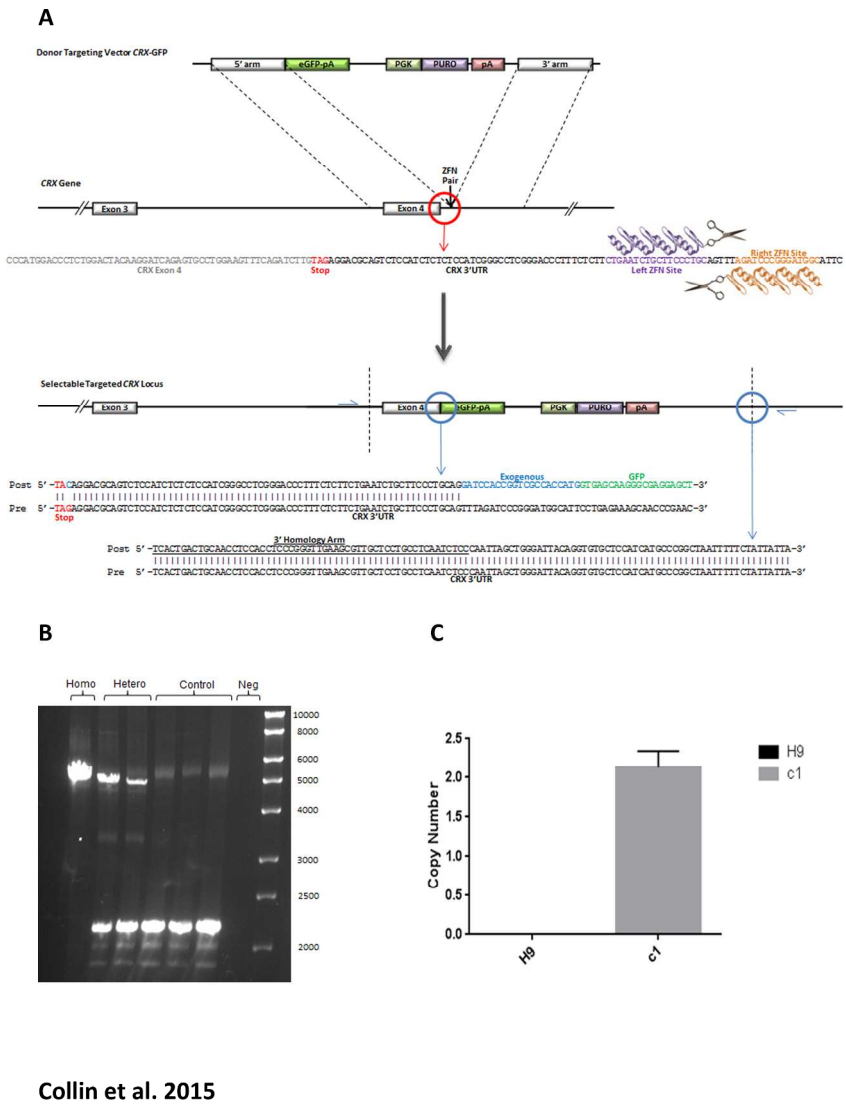
Suppl. Figure 5: CRX and RECOVERIN expression during early human fetal development at 16 (A) and 18 weeks of gestation (B-D). Double positive CRX and RECOVERIN expressing cells are located in the developing outer nuclear layer. In addition CRX+RECOVERIN- cells are found in the neuroblastic layer. Scale bars= 100 µM (A) and 50 µM (B-D).

Suppl. Table 1: ZFN pair target sites, PCR primers, sequencing primers.

Suppl. Table 2: ZFN pair off-target analysis. Off-target site sequences, primer details and sequencing.

Suppl. Table 3: Antibodies used during this study.

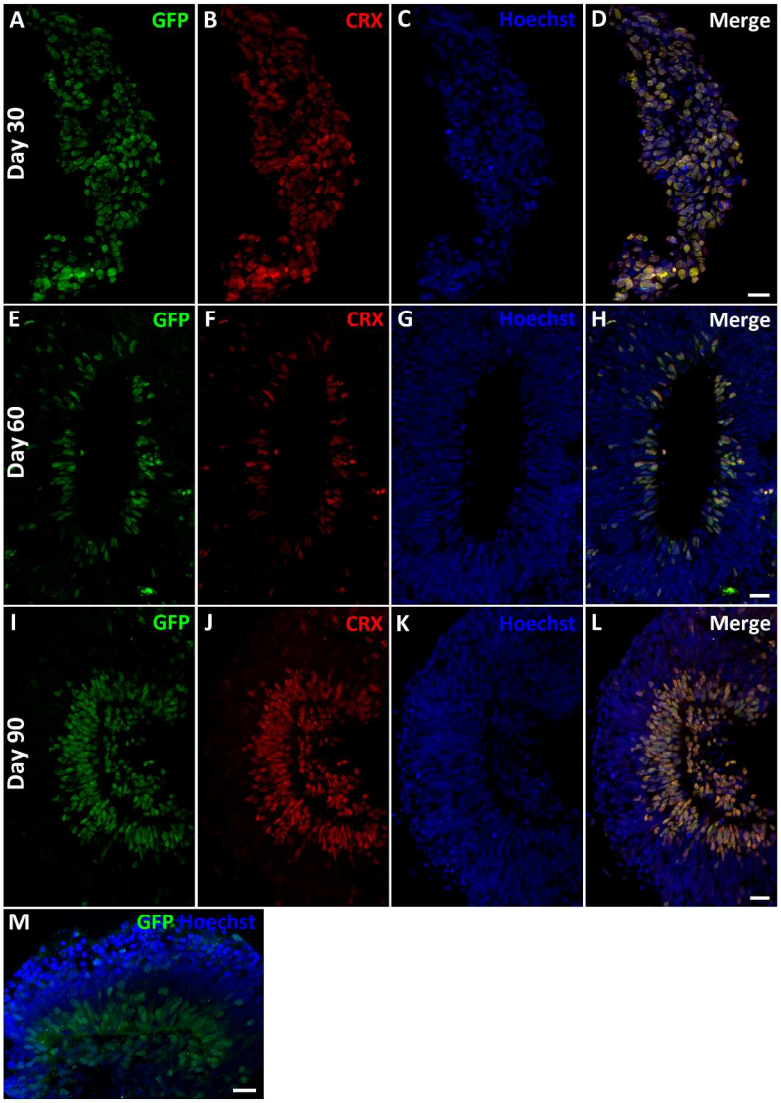
Figure 1



Collin et al. 2015

ZFN mediated targeting of GFP to the 3' UTR of CRX. A) Schematic diagram of endogenous CRX gene structure, targeting cassette, targeting strategy and sequencing data from sections of the insertion site, the endogenous CRX-exogenous eGFP border site and 3' homology arm-endogenous CRX border, obtained from three analysed hESC clones; B) PCR of three hESC clones with primers spanning the integration site to assess whether the cassette had integrated at the correct site. Depending on size of band produced (upper band of 5067bp with integration, lower band of 2171bp without) and whether the clone has a heterozygous integration (by the presence of the upper and lower bands), or a homozygote integration (a single upper band), such as clone 1 in the first lane. Lanes 2 & 3 show two clones with heterozygote integration, lanes 4-6 are genomic DNA from control untreated H9 hESCs and lane 7, a no template negative PCR control. This is a representative example of at least 3 repeats; C) qPCR mediated copy number analysis indicating presence of two copies of GFP (and hence targeting cassette) in CRX-GFP hESC clone 1 (c1) and its absence from wild type H9 hESC genomic DNA, data is presented as the mean ± SEM (n=3).
209x297mm (300 x 300 DPI)

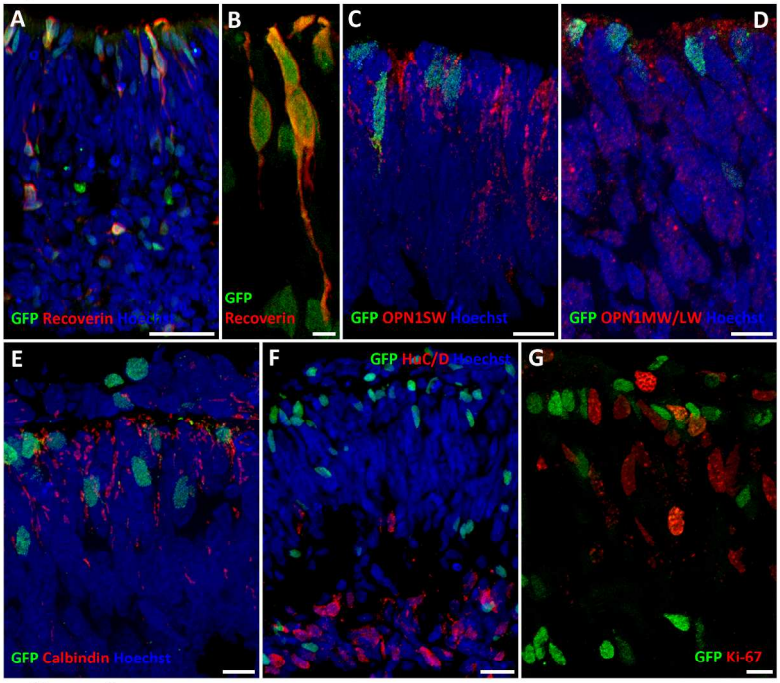
Figure 2



Collin et al. 2015

GFP reporter accurately mimics the expression of endogenous CRX during hESC differentiation. (A-L) Immunocytochemistry with antibodies raised against CRX and GFP showing correlation at day 30 (A-D), day 60 (E-H) and day 90 (I-L) of hESC differentiation. M) Detection of GFP without antibody using confocal microscopy. Scale bars= 20 µm.
209x297mm (300 x 300 DPI)

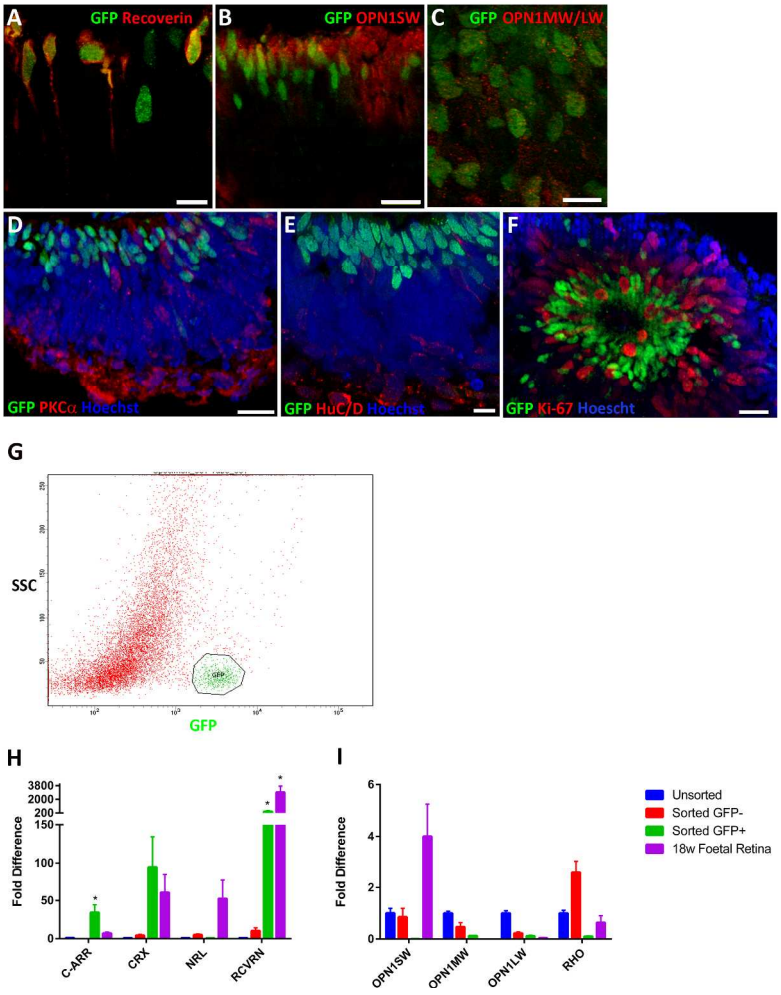
Figure 3



Collin et al. 2015

CRX/GFP expression is found in photoreceptor precursors at day 60 of hESC differentiation. Immunocytochemistry with antibodies raised against GFP and Recoverin (A and B), OPN1SW (C), OPN1MW/LW (D), Calbindin (E), HuC/D (F) and Ki-67 (G). Scale bars = 20 μm (A, F), 5 μm (B) and 10 μm (C, D and E).
209x297mm (300 x 300 DPI)

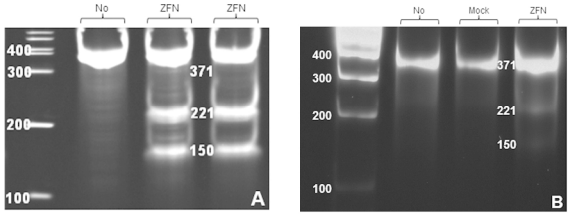
Figure 4



Collin et al. 2015

CRX expression in photoreceptor precursors assessed by immunocytochemistry and qRT-PCR at day 90 of differentiation. Immunocytochemistry with antibodies raised against GFP and Recoverin (A), OPN1SW (B), OPN1MW/LW (C), PKCα (D), HuC/D (E) and Ki-67 (F). Scale bars = 10 μm (E) and 20 μm (A, B, C, D and F); (G) Flow activated cell sorting of GFP positive (green) and GFP negative (red) populations; (H) and (I) qRT-PCR analysis of unsorted, flow activated cell sorted GFP positive (Sorted GFP+) and GFP negative (Sorted GFP-) populations and foetal retina at 18 weeks of gestation (18w Foetal Retina), indicating CRX, RECOVERIN (RCVRN) and CONE ARRESTIN (C-ARR) expression in the GFP positive population. Expression of short, medium and long wave opsins (OPN1SW, OPN1MW & OPN1LW) and rhodopsin (RHO) are also shown. Data is presented as the mean ± SEM (n=3). Significant differences found are marked with an asterisk (*).
209x297mm (300 x 300 DPI)

Suppl. Figure 1



Collin et al. 2015

Surveyor mutation detection assay performed in K562 cells and hESCs to indicate ZFN cutting efficiency. As shown, a higher ZFN cutting efficiency is observed in K562 cells than hESCs (which may be related to lower non homologous end joining repair pathway efficiency observed in hESCs). Upon cleavage of the target site by the ZFN pair inaccurate repair by the non-homologous end joining pathway can occur resulting in insertion or deletion of nucleotides. Following PCR of the target region and cooling of the PCR products duplexes form. In the presence of NHEJ-repaired and wild-type heteroduplexes and the SURVEYOR/CEL-I enzyme cleavage products are produced. The 221 and 150 bp bands represent the cleavage of PCR products generated across the CRX 3' UTR target cut site in the presence of an insertion or deletion, in the absence of an insertion or deletion a wild-type PCR product of 371 bp is observed. (A) K562 cells; lane 1, PCR of wild-type genomic DNA (No), lanes 2 & 3, PCR of genomic DNA following ZFN pair transfection (ZFN); (B) hESC H9 cells; lane 1, PCR of wild-type genomic DNA (No), lane 2, PCR of genomic DNA following a mock nucleofection without ZFNs (Mock) & lane 3, PCR of genomic DNA following ZFN pair nucleofection (ZFN).
209x297mm (300 x 300 DPI)

Supplementary Figure 2: CRX Donor Cassette Sequence

GTGTCCTCATCCCCGGGCACCCTGCGGCTCTCCTGGGCCTCTTCCCCACTTACCCACCCCCATCTCCGCTCTTATCCCCCAGGTTTGGTTCAAG
AACCGGAGGGCTAAATGCAGGCAGCAGCGACAGCAGCAGAAACAGCAGCAGCAGCCCCCAGGGGGCCAGGCCAAGGCCCGGCCTGCCAAGAGGA
AGGCGGGCACGTCCCCAAGACCCTCCACAGATGTGTGTCCAGACCCTCTGGGCATCTCAGATTCCTACAGTCCCCCTCTGCCCGGCCCCTCAGG
CTCCCCAACCACGGCAGTGGCCACTGTGTCCATCTGGAGCCCAGCCTCAGAGTCCCCTTTGCCTGAGGCGCAGCGGGCTGGGCTGGTGGCCTCA
GGGCCGTCTCTGACCTCCGCCCCCTATGCCATGACCTACGCCCCGGCCTCCGCTTTCTGCTCTTCCCCCTCCGCCTATGGGTCTCCGAGCTCCT
ATTTCAGCGGCCTAGACCCCTACCTTTCTCCCATGGTGCCCCAGCTAGGGGGCCCGGCTCTTAGCCCCCTCTCTGGCCCCTCCGTGGGACCTTC
CCTGGCCCAGTCCCCCACCTCCCTATCAGGCCAGAGCTATGGCGCCTACAGCCCCGTGGATAGCTTGGAATTCAAGGACCCCACGGGCACCTGG
AAATTCACCTACAATCCCATGGACCCTCTGGACTACAAGGATCAGAGTGCCTGGAAGTTTCAGATCTTGTACAGGACGCAGTCTCCATCTCTCT
CCATCGGGCCTCGGGACCCTTTCTCTTCTGAATCTGCTTCCCTGCAGgatccaccggtcgccaccatggtgagcaagggcgaggagctgttcac
cggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggc
aagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttca
gccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacga
cggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggc
aacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttca
agatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgctgctgcccga
caaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccggg
atcactctcggcatggacgagctgtacaagtaaagcggccgcgactctagatcataatcagccataccacatttgtagaggttttacttgctttt
aaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaaa
taaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttaagggt
accgagctcgaattccgatcatattcaataacccttaatataacttcgtataatgtatgctatacgaagttattaggtctgaagaggagtttac
gtccagccaagcttaggatctcgacctcgaaattctaccgggtaggggaggcgcttttcccaaggcagtctggagcatgcgctttagcagcccc
gctggcacttggcgctacacaagtggcctctggcctcgcacacattccacatccaccggtagcgccaaccggctccgttctttggtggccccctt
cgcgccaccttctactcctcccctagtcaggaagttcccccccgccccgcagctcgcgtcgtgcaggacgtgacaaatggaagtagcacgtctc
actagtctcgtgcagatggacagcaccgctgagcaatggaagcgggtaggcctttggggcagcggccaatagcagctttgctccttcgctttct
gggctcagcagctgggaagggtgggtccgggggcgggctcaggggcgggctcaggggcggggcgggcgcccgaaggtcctccggaggcccggca
ttctgcacgcttcaaaagcgcacgtctgccgcgctgttctcctcttcctcatctccgggcctttcgacctgcatccatctagatctcgagcagc
tgaagcttaccatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccagggccgtacgcaccctcgccgccgcgttcgc
cgactaccccgccacgcgccacaccgtcgatccggaccgccacatcgagcgggtcaccgagctgcaagaactcttcctcacgcgcgtcgggctc
gacatcggcaaggtgtgggtcgcggacgacggcgccgcggtggcggtctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgaga
tcggcccgcgcatggccgagttgagcggttcccggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggcccaaggagcccgc
gtggttcctggccaccgtcggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgc
gccggggtgcccgccttcctggagacctccgcgccccgcaacctcccctttctacgagcggctcggcttcaccgtcaccgccgacgtcgaggtgc
ccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgcctgacgcccgccccacgacccgcagcgcccgaccgaaaggagcgcacgacc
ccatgcatcgatgatatcagatccccgggatgcagaaattgatgatctattaaacaataaagatgtccactaaaatggaagtttttcctgtcat
actttgttaagaagggtgagaacagagtacctacattttgaatggaaggattggagctacgggggtgggggtggggtgggattagataaatgcc
tgctctttactgaaggctctttactattgctttatgataatgtttcatagttggatatcataatttaaacaagcaaaaccaaattaagggccag
ctcattcctcccactcatgatctatagatctatagatctctcgtgggatcattgttttctcttgattcccactttgtggttctaagtactgtg
gtttccaaatgtgtcagtttcatagcctgaagaacgagatcagcagcctctgttccacatacacttcattctcagtattgttttgccaagttct
aattccatcagaagctggtcgagatccggaacccttaatataacttcgtataatgtatgctatacgaagttattaggtccctcgaagaggttca
ctagtactggccattgcggccatcTTTAGATCCCGGGATGGCATTCCTGAGAAAGCAACCCGAACCAGCTGTCCTTCTGACAGCTCGGTGTTCA
GCTTACAGAGACCACCCCTTTCCTCCACAGGGAGAGGCTCCTCCCTCTCCTGGGACAGCTCACAGGTCCTAGTGATTCTCTCAACCCTAACACC
GTCTGGCACGATTGTGACCGCTGAAGTACACCACGAGCTCCAGGCTTCAGAAAGTGGTGCTGAGAACTTGCTCCAAGAAGAAGTCAAACCAAAC
TTGCAGTTGATTTGGGGTCATGTTTAGGTCAGAATCACCGTGCCCTTGAACAAGCAGGTAGGGGGGCTTGATAACTTAACTTTCCACGTGGACA
GAATTTTTTTTTTTGTTTTGTTTTTGTTTTGCAGACACAGTCTAGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCACGATCTCAGCTCACTGCAA
GCTCTACCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGATGCCCACCACCAGGCCCGGCTAATTTTTTTT
GTATTTTTAGTAGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCCGCCTCGGCCTCCCAAAGTGCT
AGGATTACAGGCGTGAGCCACCGCGCCCGGCCCTTTTTTTTTTTTTTTTTTTAATTGAGACGGAGTCTCACTCTTTTGCTTAGGCTGGAGTGC
AGTGGTGTGATCTCAGCTCACTGACTGCAACCTCCACCTCCCGGGTTGAAGCGTTGCTCCTGCCTCAATCTCC

Key:

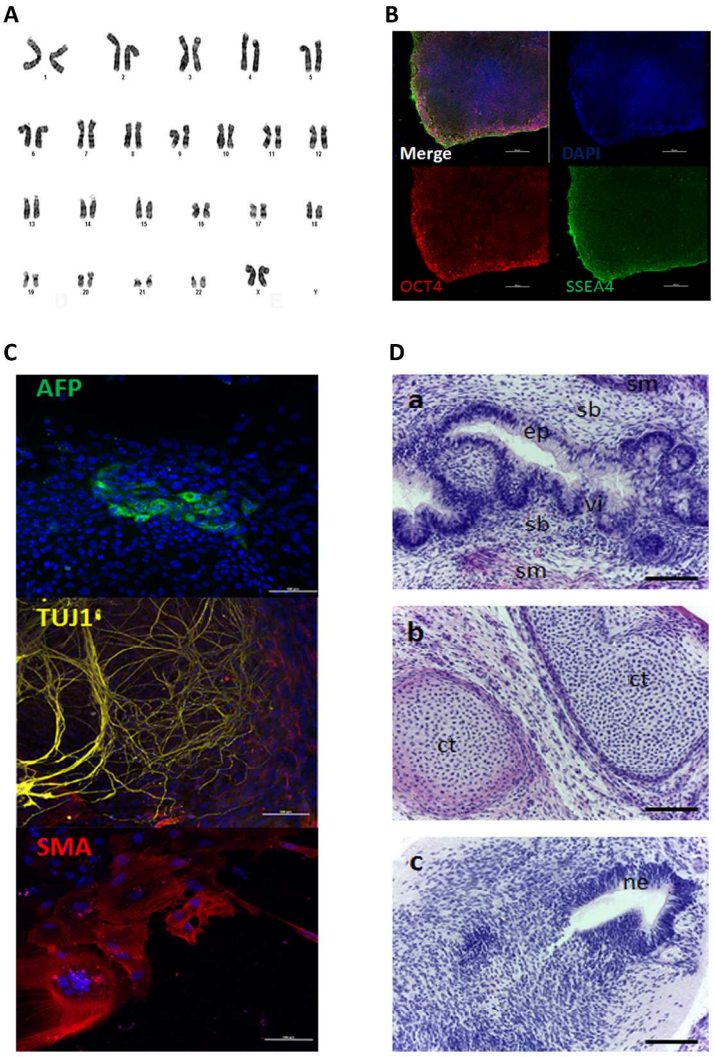
5’ and 3’ CRX Homology Arms

eGFP

PGK Promoter

Puromycin

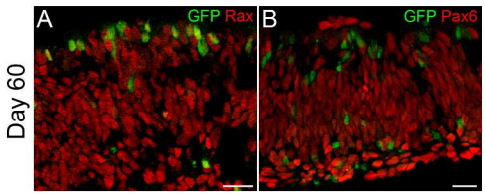
Suppl. Figure 3



Collin et al. 2015

CRX-GFP targeted hESC clones maintain a normal karyotype and exhibit pluripotency. (A) Normal karyotype (46 XX) observed in CRX-GFP hESC clone 1; (B) Pluripotent marker staining (OCT4 and SSEA4) in CRX-GFP hESC clone 1; (C) Three germ layer differentiation capacity assessed by the presence of endodermal cells (marked by AFP staining), ectodermal cells (marked by TUJ1 staining) and mesodermal cells (marked by SMA) in embryoid bodies derived from differentiation of CRX-GFP hESC clone 1; (D) Histological analysis of xenograft tumours formed from engrafted CRX-GFP hESC clone 3. The teratomae produced contained tissues representative of endoderm (a), mesoderm (b) and ectoderm (c). Example tissues included: (a) structure of primitive intestine showing epithelium (ep), villous (vi), submucosa (sb), and smooth muscle (sm); (b) cartilaginous masses (ct); (c) neuro-epithelium (ne). Counterstain: Haematoxylin and Eosin. Scale bars: (a-c) 150 microns.
209x297mm (300 x 300 DPI)

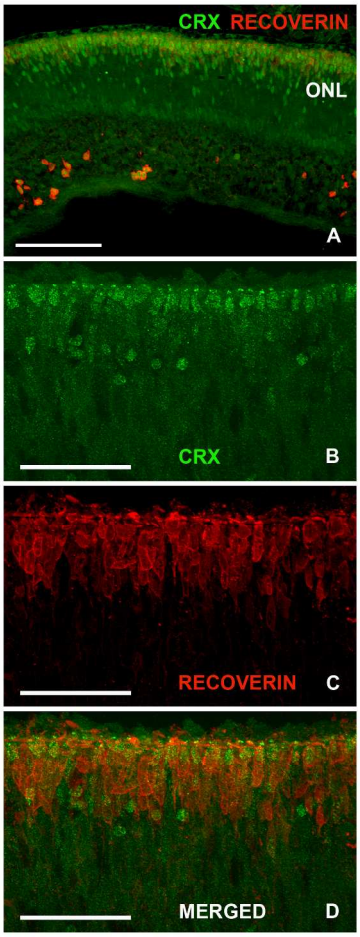
Suppl. Figure 4



Collin et al. 2015

CRX is not expressed in early stages of eye cup formation. Immunocytochemistry with antibodies raised against GFP, RAX (A) and PAX6 (B) (both markers of eyefield development) at day 60 of differentiation. Scale bars= 20 μM.
209x297mm (300 x 300 DPI)

Suppl. Figure 5



Collin et al. 2015

CRX and RECOVERIN expression during early human fetal development at 16 (A) and 18 weeks of gestation (B-D). Double positive CRX and RECOVERIN expressing cells are located in the developing outer nuclear layer. In addition CRX+RECOVERIN- cells are found in the neuroblastic layer. Scale bars= 100 μM (A) and 50 μM (B-D).
209x297mm (300 x 300 DPI)

Supplementary Table 1: ZFN target sites and primers used in this study

ZFN Target Sites	Left	CTGAATCTGCTTCCCTGC
	Right	AGATCCCGGGATGGC
Primers Flanking ZFN Target Site	Forward	GGCACCTGGAAATTCACCTA
	Reverse	CCACTTTCTGAAGCCTGGAG
Primers Flanking Integration Site	Forward	CCCACAGCTGGATGCAAAGT
	Reverse	AGATGGAACAGGCAAGGTGC
GFP Sequencing Primer		TTACGTCGCCGTCCAGCTC
Puromycin Sequencing Primer		GCATGGCCGAGTTGAGCGGT
Copy Number TaqMan Assays	EGFP	Mr00660654_cn
	RNase P	4403326
Gene Expression TaqMan Assays	CRX	Hs00230899_m1
	EGFP	Mr04329676_mr
	RHO	Hs00892431_m1
	NRL	Hs00172997_m1
	RCVRN	Hs00610056_m1
	ARR3	Hs01020134_m1
	OPN1SW	Hs00181790_m1
	OPN1MW	Hs04194752_g1
	OPN1LW	Hs01912094_s1
	GAPDH	Hs99999905_m1
	MRPL19	Hs00608519_m1
	RPLP0	Hs99999902_m1

Collin et al. 2015

Supplementary Table 2: Off-target site analysis performed with PROGNOS and sequencing of the resu

PROGNOS (Version 1) Ranked Output:

Ranking	Nuclease Orientation	Spacer Length	Mismatches	Left Half Site	Spacer
1	L-5-R	5	0_0	CTGAATCTGCTTCCCTGC	agttt
2	L-6-L	6	4_4	CTacgcCTGCTTCCCTGC	gatcca
3	R-5-L	5	4_3	GCCATgagtGGATCT	gaagc
4	R-6-R	6	5_1	cCCATCCCGGGATCT	ccctgt
5	R-6-L	6	5_3	aCCATCCCaGGcTCT	cccgca
6	R-6-L	6	5_5	GCCtTCCCtcccTCT	gctgct
7	L-6-L	6	5_5	CTGgAcCTGaTaCCCaGC	aggtgt
8	L-5-R	5	5_5	accAAgCTGCTgCCCTGC	acagg
9	L-5-R	5	4_4	CTGgcTCcaCTTCCCTGC	ttttt
10	R-6-R	6	5_3	GagATCagaGGATCT	ccaaac
11	L-5-L	5	5_5	gccAAaCTGCTTCCaTGC	aggag
12	R-5-R	5	4_4	cCCcTCCCtGGATCa	tgtct
13	L-5-R	5	5_3	CTGcAgCTGCTTCCCaGC	acaaa
14	R-6-R	6	5_2	GCCcTCCtGtGAaCc	cagcac
15	R-6-L	6	5_4	GCCATCtgGGttTCT	gccagg
16	R-6-R	6	5_4	cCCAaCCCtaGgTCT	ccacaa

Collin et al. 2015

ıltant PCR products from H9 CRX-GFP clone 1 genomic DNA. Ranking number 1 is the correct ZFN pair

Right Half Site	Chromosome Name	Chromosome Location	Genomic Region
AGATCCCGGGATGGC	chr19	48343274	Exon
GCAGtGAAGCAGgTgtAG	chr8	74689849	Intergenic
GCAGGGAgGCAGAagCAG	chr7	26661904	Intergenic
AccTCCtGGcAgGGC	chr19	6590640	Intron
GgAGGGAAGCAGcTcCct	chr6	160308315	Intergenic
GCAGGGAAGCtGAccaAa	chrX	48460428	Intron
GCAGGGAAaggtATgCAG	chr19	31560623	Intergenic
tGgTCCCGGcgaGGC	chr1	10748837	Intron
AaATCCtGGGATtGa	chr6	168582079	Intergenic
AGATCCaGaGAaGGC	chr1	40954614	Intron
GgAGGGAAGCctgcTCAG	chr17	47975818	Intergenic
AGAagtCaGGATGGC	chr11	5405799	Intron
cGgTCCtGcGAgGGC	chr18	76007140	Intergenic
AGATCCtGGGcTGGC	chr16	86016582	Intergenic
GCAGGGAAaCcaAagCAG	chr2	62597922	Intergenic
AGActtCGGGATGcC	chr17_gl000206_random	37198	Intergenic

targeting sites.

Closest Gene	Forward Primer	Reverse Primer
CRX	CCATGGACCCTCTGGACTAC	GCTCGTGGTGTACTTCAGCG
UBE2W	CTCTTTCTAGAAAGCCCTCTCCTC	GGACTCCATTTCAGAGGGCATGT
C7orf71	ACAAACCTGCACATTGACCCCTG	CACTATGCCTGGCCAAAACCTG
CD70	TCTCTACGCTGCAAAGGCGC	TGGGTACAAAGAGGCCAGAGAG
MAS1	GCCTCACAGCATGCTCGGTT	TCATCTGGCACAAGTGCCTTTGCG
WDR13	TGACAGGCCGTGTCCTTGCT	GAGGCAAGCATTGATGAGCAGTG
DKFZp566F0947	ATTCTCTACTGTGGCATGCTCCTG	CATGGATGGCAGTCCCATGG
CASZ1	AGGCAGACGCTGCCAACACA	CGGCAAGCAGCAGCTGTTCT
FRMD1	GGATCCCATGGATATGAGGGGC	AGAAAAGCAGCCCCAGGGAC
ZNF642	CACTGTGGTTCTCCCAATGTCTC	CCACTGACACCAAGTTGCTGTAG
FLJ45513	GCAAACCCAAGCCAGCTCCT	GTCAAATCCCCCACTACAGAGACA
OR51B5	CTTCTCTTGCTGGAAAAGAGGCG	GGATACTGGGGAAGTTACAAGGTG
SALL3	AGGGTCTCACACGTGCTGCT	TTTCATGATGCCGCCCATGTGAC
IRF8	TATGTGTGGCCATGCGTGTGTG	ACTGTGTGGCCTCAGGCAAAC
TMEM17	GTAGTTGTGGATCCATTCTCCAGC	GCACTTCTACCAAATCCCTTTGCC
N/A	CTCACAGTCTGTCCTGACCC	GCCGAGGCTTCTTTCACTCTTC

Number of indels found
N/A
None
None
None
None
None
None
None
None
None
None
None
None
None
None
None

PCR Product Sequence
N/A
CCCTCTCCTCTTTAGTACTCCCACCTGAGGACTCCAGTCACCCTGGGCCTCCCTGGACTTCCGGCTTTGTCTCCCCAGGT
TTGACCCCTGAACCTAAAATTAAAAAGAAAAAAGAATAAATAAATAAAACTGGCTTTCTAAATATGCTCTGGAAGAGCTG
AAGGCGCCCGGCGCGGTCCCCAAAGGGACCCCTCCCCCACCCGGAGCGAGTGGCTGGGGCCATCCCGTCCTAGGAGGCCC
GCTCGGTTTATTTTCTAATTATTTTCTCCTGCTGCAAACTTTAGCCAGCCCTGTAGATATGATATCTGGTGATTCTGGGG
TCCTTGCTCTGTCCTTTGATGCCCCTGGCCGGCTGCTCTGGGCGGGTGATGACCATGGCAGTGTCTTCTCTTTCCTCTTT
GCATGCTCCTGATGAAATTCAACAGTTGGGACAAATGGGCTAGAAAGAAACCTCTCTGAGGGCAGAGGCCTTAGGTGCCA
CCAACACAGGGGCCCGGGCAGTCCTCATTTCCGTCCCTCTTGCTGCAGGAGAGCCTCCCATGCCTCCTGGGTTTCAGGCC
ATGAGGGGCCAAGCTATGACTTAGAGAGTTTTGGACTTAGAAGCCTGGGTTTTCCCTGTATCTCTCTATGTCCTGGGGCT
CCAATGTCTCTTATCCTGCTTAATTTTTTTCTAAGGTATGTATCACCATCTGTCTTTCTATATATTACATTATTTATAGT
CAGCTCCTGGGGTCCTACAGGGAGGGAGGGCATGTAAATCTTGCCTCACCCTCTTCCCTCCCAGCCCTGCTGTCAGCTGG
GAAAAGAGGCGAGCTCATCTGCAGTATTCTGTTAGGAAGGCCAGGAGGCAGGGCTGCCTCACTCCCCTGATGAGTGCCTG
GTGCTGCTGGGGAGGGGGTGAGGGGGGAGGACCACCACGGAGAGCTTAGAGCTGCACATAGGACTCGGGTGGGTGTCAGG
CCATGCGTGTGTGAGTGTGGCCTCTCGTGTCAGTGGGTGGGCCTGCTGTCGGCTCCTGCAGTCCCCTGTGGCCTCTTCTC
TCCATTCTCCAGCTTGGGAATTTCCAGCCTCTGTGATTTTCTCTCCCCCATCCCCACACCCGCACCTCCAGGGCTAGGCT
TCCTGACCCTGGCTCTGTGGCCCCTCCTTCTCCTTGATCAGGCCCCTCCCGGGTCTACCGGGATCTCTGGCTGCCCCCCG

```
CAGGAGGATGGCTGGGCTACGCCTGCTTCCCTGCGATCCAGCAGTGAAGCAGGTGTAGGGCTCACTCCATCTGT
ATGAAAATGACATCTCTTAGGATGCAATTTGGAAAGAAGTGACCTGCACCTTCCCAAGGCTTCCTTTGTCCCAC
CAGGAAGCCCCCACGCCCAGCCTTGTGAAGCCCCCAGCAGCCTCTGAGTCAGCCTGCCGGGGGAACACCAGAGC
GACTCAGAGAGGCATTTATTTATTTTCTTATTTATATGAGCTATATTAGTTGGGAGTTCTCCTTTTGGAACTAG
GATATGGCCACAGGTAGGCAGACAGCAGGCCTGCATCTGGGTGCTCGCCCTCCAGCCCTCCCAGGGCACAGCCC
CAATGCTTACCTGGACCTGATACCCAGCAGGTGTGCAGGGAAAGGTATGCAGAAGAGAAGGAAGGGGATGGCAT
TCGCCTCCCACATGGGTGTCTTACATCCTGCTGAGTCGAGGGTGTCAGGGTGCTGTTCTGTGGCCAGGCAGCCA
GGGGAAGTCCCATGGGAGCAGGATTCTCAGACCTCAGGCCATGGGGACCCGTGCTGTGGACCCCAGGGATGTGA
AATCTCCAGATTAGAGATCCAGACAGACTACAGTTCCAGACAGTTTTGTTGATCAGATTTGGAGATCAGAGGAT
CCAGCCCCATCCCTGGAGCCAAACTGCTTCCATGCAGGAGGGAGGGAAGCCTGCTCAGTGGGAAAAGATGAAGC
CATGGCCAGGCCAGCCCCTCCCTGGATCATGTCTAGAAGTCAGGATGGCAGAAGCCCTCTGGTCTCTATACTCT
GTGAGATTTCTGAGATAATGAGTCTCATTACACCTCCGAGGTTGGCTACCCAGAAACCCAGCTACCTTAGGGGC
CAAGGACAGGTAGCAGAGCCCTCCTGTGAACCCAGCACAGATCCTGGGCTGGCTGTGGTCCGGGAGGGCCAGAG
TTTGGCTGCTTTGTGGCCATCTGGGTTTCTGCCAGGGCAGGGAAACCAAAGCAGATATTTACTAAGAGCTAAGC
ACCCCCTGCTGCAGCCCCAACCCTAGGTCTCCACAAAGACTTCGGGATGCCCCGTGAGCCCCTGGGTCCTGGAC
```

:TTTCCTTCGTTCAGTGACTGATGCCCTTTTTTGCTTGCTAGCCAGTGTTTTGAAAATTATTGTTTCATAT
:TCCCACGGCAGTGAGCATGTTTCACATGCCATGAGTGGATCTGAAGCGCAGGGAGGCAGAAGCAGCCACC
:CGAGTCATCTTCCATTTCCATGTGTGCCTACCTTTCCCATCCCGGGATCTCCCTGTACCTCCTGGCAGGG
GTGAGGTCTAGTTTGTCCACTGCGATGATCTCTGGGGCAGCGGACACCATCCCAGGCTCTCCCGCAGGAGG
:TCACCCTCACCTTTGCCTTCCCTCCCTCTGCTGCTGCAGGGAAGCTGACCAAAGCCAAGCGTTTGGTGGT
:GGCATGCACAGCGGCAGCACACAGGCTGTGTGTGGGCAGTGCCCAGGAAGGAGGTGCGTCAATACAACAA
AGAGGGGTGTGCGCTAGACCACCAAGCTGCTGCCCTGCACAGGTGGTCCCGGCGAGGCTGCCCCTCCTCTT
ATGTCCCTGGCTGTTCATTTCCACCCTGGCTCCACTTCCCTGCTTTTTAAATCCTGGGATTGAGGGAACAT
:CTCCAAACAGATCCAGAGAAGGCTCCTGATAAAAGTAAGTTTAGAATCTCTCAGAATTTAGAATACTGTC
:TGGTCCCCCCGCGTGTCAGTGAACAAATTGGGCCAATTTGGTGTGATTAGTGAAGAATTCACATCTGGCC
:TTCCTGCCGAGCCATCTGGTATACACACCCTCTCCTTACTCTCATCAGAAAAACAGTCATTCACTTATTC
GGTAAATATAATGTAAAGAGGAGAGGAGACTAGTGCGGGGAAATGAAGGCTGCAGCTGCTTCCCAGCACAA
GCATAAGGGAATAAGTGTGGAGCGCTTGCTCTAGGCCCAGCCCTGTGCTAAGCAGACCATAGCTGTAACCC
:AGGGCAGGGGTTTCTCCACTCCACACAATAGGTACAGCACTAGGGCCTACAAAAGTGTTTAAGGCCTATG
:CCCTCTTCCTCCTCCAGCCCCTCCTCTCCCAGCTGCTCCTGGCTGCCCTTTAATTGCTCAGAGCTGTTCA

'ATTTCATCTGAGTTTTTAGTTCTTTCAGGCTTTTTAGTTCTTTAAATCCAGTCCCCGT
'GTTTTCTGGAACCACCTGAGTGGCAGTCATGGGGGTCCATTTGTGGAAGAAAGGAAATGTTTCCAGGCAG
'CTGCCTGTCTCGTAGCCCCTACACCGGGCAGCCTGGTCCCCGCCTCGCCTCCCTGTTCTGGTCTCTGTCT
'GAAGCAGCTCCCTGAGCCTCGCCCCGCCCCTCAGTGGAAGAGCTATCCTGGATTAGGGATGCCCGACTCT
'GCATGAGGGGAGCCCTGTGACCAGCATCTCAGCCCGGTCCTGGGTCAGCCGCGAGGCCCGGGATCCCTCA
'GGGTTGACACCAAGTGCAGTGCAGTTGAAAGCCACAATGTAATTATCCACCCAGGAAAAGAGGGTCCCCT
'TTGCACACTGGGGCTGCCCGCCCTCCTCCTGCTCCCAGCTTCTCCCTACTCATTGTACCCTGGCCCAGCT
'ATGTCAAAGACGTGGCACATGTGTGGCTGTGGGGCAGCAGCCCCTGGAGAGAGGAACGCCCCCAGGCCTG
'AGAACTCCTTGGAACTTAACAGAGGTGGTCTCCGGCTAAAAATGAATGTATTTGATGTTCTAGGAATTGT
'AGCAGAGTCCTCCCCCCAGCCCCAGCCCAGCCTGGCTCCTGCTCCCCATTAGAAGCTATTTTCTCTGACG
'AACAAATTATTAAATGCATTTTATATGCTGGTCCAGTTAACATAGAAAAGACCCCATCCTCAATCTGCTT
'ACGGTCCTGCGAGGGCACCCGCACCCTCCCTCCAGAACGCAGCCGCTCTCCCGGGGACGGAAGGAAAGTT
'AAGACACTATGAGATAAGAACTACTTTTGGCCCTATCGTACAGAGTCTGAGATGAATTCAGAGAGGTCGA
'TGGCAGGCAGAAAAATGCTCCTGCAAAGATGTGCATGTCCTAATCCCCCAAACCTGTGAATATGTTACCT
'GCAACCTGGCGCGTAGCTGGTACATAAGTGGCTGTTGGTTGGAGTCAAGTGAGTGATGGGGAGATGCTAA

ACATTTCATGGTGGATTCAAAGATTACAGGTT
CTGCCCTACCAGATCTTCCTCTCTGGCC
GGTCACTTCACATCCTCCTCACTGCTCCGCAAAGGC
CTGCTCA
CCCATGC
ACCCCAGAACAGCTGCTGCTTGC
TGCTCCTGGGCTCCCACCTCGGGTCCC
TGACTTTCAAGGACATATCTATTGACTTCACCCAGGAAGAGTGGGGGCAGCTGGCTCCTGCTCACCAGAA
GCTCCATTTGTCTCTGTAGTGGGGGATTTC
ATTTTAATCCAAACAACACCTTGTAACTTCC
CAAGCACCGGGGCCCTCGCAGTGGGAGACCAGTCACATGGG
GTTGTTTGCCTG
TACATGGCAAAGGGA
GAAGAGTGA

TCTATACCGAGAGGTGATGCTGGAGAACTACAGC